DOCUMENT RESUME

ED 093 906 TM 003 721

AUTHOR Taylor, Bob L.
TITLE Potential Uses of the National Assessment Model at the State and Local Levels.
PUB DATE [Apr 74]
NOTE 31p.; Paper presented at the Annual Meeting of the American Educational Research Association (59th, Chicago, Illinois, April 1974)

EDRS PRICE MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS Academic Achievement; *Citizenship; Curriculum Development; Educational Accountability; *Educational Assessment; Information Dissemination; *Models; National Surveys; Objectives; *School Districts; *State Programs; Testing Programs; Use Studies
IDENTIFIERS *National Assessment

ABSTRACT

The model used by National Assessment for data gathering and reporting on the Citizenship area is described, and the potential uses of the model for state and local assessment, curriculum development, and accountability purposes are discussed. The study was carried out using papers and reports from the National Assessment of Educational Progress, Denver office, and state reports on adaptations of the model for state assessment needs. Adaptations of the model for curriculum development were identified, and, finally, adaptations of the model for accountability purposes were suggested and discussed. (Author)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

BEST COPY AVAILABLE

ED 093906

POTENTIAL USES OF THE NATIONAL ASSESSMENT MODEL AT THE STATE AND LOCAL LEVELS

By

Bob L. Taylor

University of Colorado

Boulder, Colorado

AERA Annual Meeting

Chicago, Illinois

April, 1974

TM 003 721

BEST COPY AVAILABLE

POTE"TIAL USES OF THE "ATIONAL ASSESSMENT MODEL AT THE STATE AND LOCAL LEVELS

by

Bob L. Taylor
University of Colorado

National Assessment is a plan for the systematic, census-like survey of knowledges, skills, understandings, and attitudes. It is an information gathering plan aimed at providing both educators and the lay public with information concerning the level of achievement in selected subject areas for students and young adults. The goal is to provide information that can be used to improve education. It is concerned with the achievement status of four age levels in ten different subject areas. The subject areas selected for assessment were: Art, Career and Occupational Development, Citizenship, Literature, Mathematics, Music, Reading, Science, Social Studies, and Writing.

THE ASSESSME"T MODEL

The assessment model is in the continuous process of being refined and improved, thus only the basic components of the model are presented in Diagram I. A circular scheme is used as the best way of presenting it since in actual use its application may be initiated with any one of the components. Also, in its application, there are continual interactions between and among the various components. While theoretically the process starts with the refinement of overall national goals into specific subject matter, behavioral objectives,

and progresses in logical sequence through to the final Utilization of the Information, in practice there is much greater freedom with respect to the utilization of the components.

The model for the Citizenship Assessment is presented here in outline form with a fairly detailed description of its components. As presented in Diagram I, there are seven basic components identified in the model: Objectives Development, Exercises Development, Sampling Plan, Administration of Exercises, Scoring and Analysis, Reporting and Dissemination, and Utilization of Information. While many of the fine points of the model are not developed in the following outline, it is described in sufficient detail to give the reader a good understanding of how the data were collected and what implications might result from these data. The number of subtopics in the model and their distribution indicate that the major efforts of National Assessment have been with the first five components. Components six and seven have been areas of controversy and, therefore, have received less attention until recently.

Outline of the Assessment Model for Citizenship[1]

I. Objectives Development[2,3,4,5]

---

[1]Womer, Frank B., What Is National Assessment? National Assessment of Educational Progress, Denver, Colo., 1970.

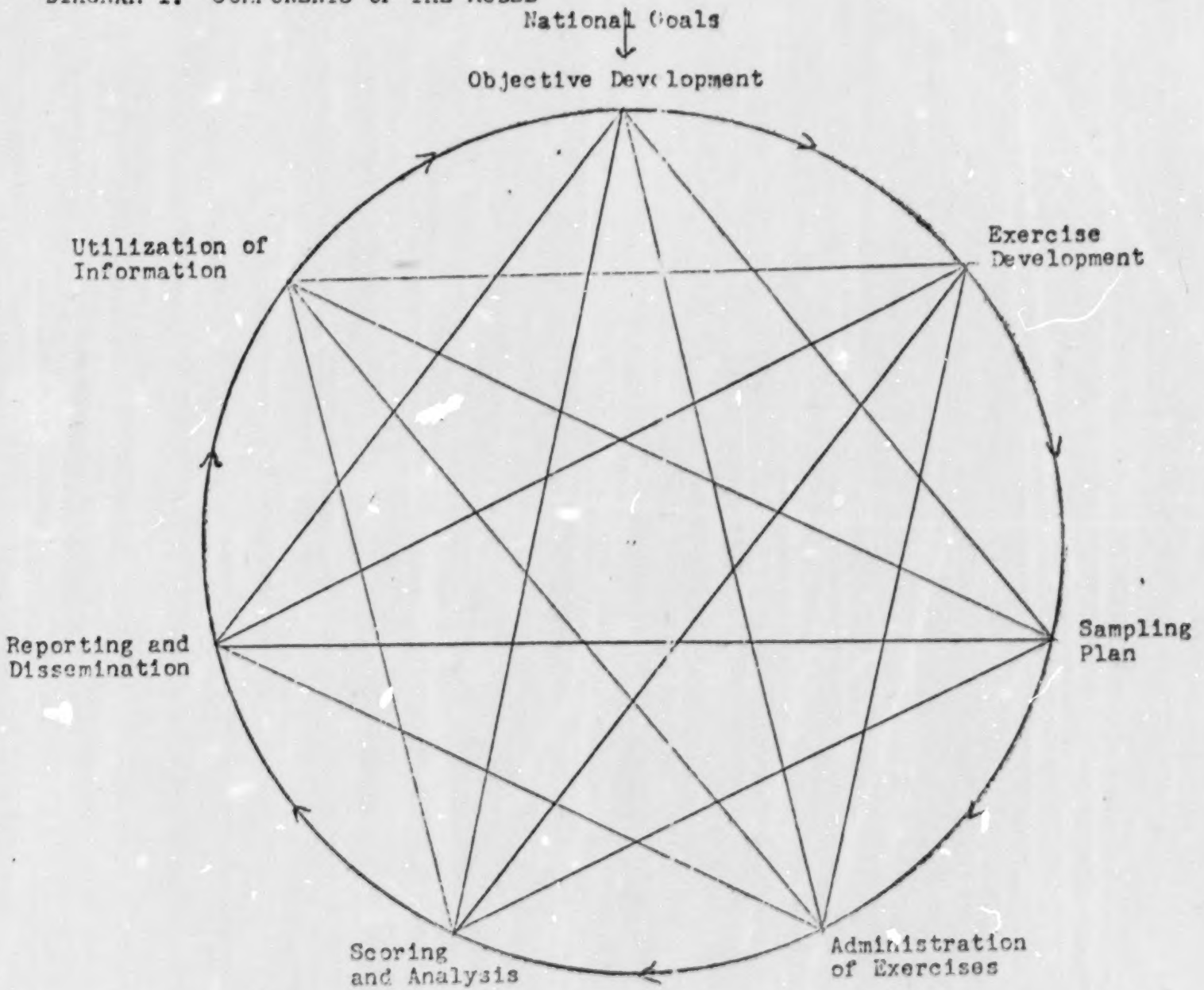
[2]Morris, Eleanor L. (Ed.), Citizenship Objectives, Committee on Assessing the Progress of Education, Ann Arbor, Mich., 1969.

[3]Campbell, Vincent N., et al, Citizenship Objectives for 1974-75 Assessment, Education Commission of the States, National Assessment of Educational Progress, Denver, Colo., 1972.

[4]Campbell, Vincent N. and Daryl G. Nichols, "National Assessment of Citizenship Education," Social Education 32:279-81, June, 1969.

[5]Campbell, Vincent N., et al, Report 2, Citizenship: National Results, Education Commission of the States, National Assessment of Educational Progress, Denver, Colo., November, 1970.

DIAGRAM I. COMPONENTS OF THE MODEL



A. The task of developing objectives in the field of citizenship was awarded to the American Institute for Research of Palo Alto, California. These criteria were used in examining the objectives:

(1) They were considered important by scholars.
(2) They were accepted as an educational task by the school.
(3) They were considered desirable by thoughtful lay citizens.

Scholars reviewed the objectives for authenticity with respect to their subject fields; school people reviewed the objectives in terms of their actual emphasis in their schools; and laymen reviewed them based on their own experiences with regard to their value in real life.

B. The American Institute for Research staff reviewed previous lists of citizenship objectives and boiled these down to one comprehensive list of 20 objectives.

C. Outstanding local teachers familiar with each target-age group (9, 13, 17, adult), working with the American Institute for Research staff, broke down each general objective into the most germane behaviors deemed appropriate as goals for a given age group.

D. A selected group of students and adults in each age group was asked by the American Institute for Research staff to recall and describe outstanding citizens of their acquaintance and specific incidents reflecting good and poor citizenship. These incidents and descriptions, about 1,000, were used to check the completeness of the initial list of objectives.

E. The objectives were stated on three levels (general objectives, sub-objectives, and behavioral age illustrations or statements). The results were summarized for each age group.

F. The revised list of objectives, broken down into important behaviors, was then deliberated on for three days by a panel of national leaders in citizenship education and related social sciences.

G. A group of persons in various roles from selected California communities reviewed the objectives and made suggestions. These included public and private school administrators, counselors, teachers, a judge, a county planner, labor and business leaders, and social scientists.

H. The objectives were then reviewed by panels of laymen. Eleven lay review panels representing four geographic areas of the country and three different community sizes were used. Each panel spent two days reviewing the objectives based on these two questions: "Is this something important for people to learn today?" and "Is this something I would like to have my children learn?"

II. Exercises Development[6,7,8]

A. The production of the exercises was initiated by the American Institute for Research in 1966. The exercises were developed to cover all of the major objectives and to represent the selected content areas. Many exercises required the use of interview techniques, as well as the usual pencil and paper exercises. Also, self report and group task exercises were used.

B. Because National Assessment intends to describe what people in an age group know, the exercises were written to reflect three difficulty levels--to report knowledge or skills common to almost all persons in an age group, to report skills or understandings of a typical member of an age group, and to report understandings or knowledge developed by the most able persons in an age group.

C. All exercises were developed to meet these criteria: content validity, clarity, functional exercise format, clustering exercises based on a single set of stimulus materials, directionality of response, difficulty level, content sampling, and overlap between age groups. The exercises were direct measures of some pieces of knowledge, understandings, attitudes, or skills which were mentioned in one or more of the objectives.

D. The exercises were reviewed by panels of lay persons for clarity, meaningfulness, and invasion of privacy.

E. There was a tryout of the exercises involving representatives of groups in the actual assessment--regions, communities, races, sexes, and age groups. Following the tryouts, the American Institute for Research staff and subject matter specialists reviewed the tryout data and made needed revisions.

---

[6]Womer, op. cit.

[7]Gadway, Charles J. (Ed.) Reading and Literature: General Information Yearbook, Education Commission of the States, Report 02-GIY, National Assessment of Educational Progress, Denver, Colo., May, 1972.

[8]Finley, Carmen J. and Frances S. Berdie, The National Assessment Approach to Exercise Development, National Assessment of Educational Progress, Ann Arbor, Mich., 1970.

F. A committee of subject matter specialists, measurement specialists, and National Assessment staff members rated the exercises to be included in the packages according to a set of criteria, and based on the ratings the exercises were selected for use.

G. The selected exercises were reviewed by U. S. Office of Education personnel for any infringement of privacy on the part of the respondents or possible offensiveness.

H. Since there were about 160 minutes of testing time available for each age group in each subject area, the exercises used were only a small sample of the potential number of exercises. The exercises were assembled into administrative units (packages) for groups up to 12 persons.

III. Sampling Plan[9,10]

A. The sampling plan was subcontracted to Research Triangle Institute, Raleigh, North Carolina. A multi-stage design was used which was stratified by region, size of community, and socio-economic status. This was a probability sample which allowed researchers to collect data from a small sample of the population and to infer from that sample certain characteristics of the entire population.

B. The populations for assessment were all 9 year olds, all 13 year olds, all 17 year olds, and all young adults 26 through 35 years old in the 50 States plus the District of Columbia. The only exceptions were the exclusions of institutionalized individuals of these ages--those in hospitals, prisons, and others who could not be reached.

C. For ages 9 and 13, a school sample only was used and for the 26 through 35 age group a household sample only was used. For the 17 year olds, both a school and a household sample were used.

D. The entire country was divided into population areas as follows: cities, counties exclusive of cities, and pseudo-counties--two or more counties were put together when the population of a single county was less than 16,000. Each population unit of 16,000 residents was assigned a number.

---

[9]Morris, Eleanor L., et al, Report 1, 1969-1970 Science: National Results and Illustrations of Group Comparisons, J. R. Chrony and D. G. Horvitz, "Structure of Sampling and Weighting," Appendix C, Education Commission of the States, National Assessment of Educational Progress, Denver, Colo., July, 1970.

[10]Morris, Citizenship Objectives, op. cit.

E. The country also was divided into four geographic regions: Northeast, Southeast, Central, and West.

F. Each geographic region was divided into communities of four types: large cities of above 180,000 population, urban fringe, middle-sized cities between 25,000 to 180,000 population, and small town-rural of under 25,000 population.

G. The 52 sampling units for each geographic area were spread across the four community types in a fashion proportional to their population in relation to the area population.

H. To insure comparable representation from each part of the country, an equal number of sampling units was selected from each geographic region--52 from each of the four regions for a total of 208.

I. Sampling units were selected at random. This plan did not guarantee that all 50 States would be included in the sample. This was not a survey objective, but later the design was changed so each state was included in the sample.

J. In each sampling unit selected, all school buildings enrolling students of the sample ages (public, private, and parochial) were identified.

K. The plan for schools was to select samples of approximately 250 to 350 pupils for each age group and from at least two different buildings within each sampling unit for each age group.

L. Each cooperating building principal provided a list of names of students in the building from the specific age groups. This list was used for the final random selection of students to take the assessment exercises from that building.

M. Information about the areas was obtained from the U. S. census data. In order to insure reliable information for lower socio-economic status groups, these groups were oversampled. There was a disproportionate number of schools from lower socio-economic status areas included. In the overall results, the data from the lower socio-economic areas were given the percentage value in which they occurred in the total population.

". From each of the 208 geographical samples, 100 adults, ages 26 through 35, were randomly selected using the following procedures. Each of the 208 geographic samples was divided into equal secondary sampling units. Then ten secondary sampling units were randomly selected from the total 208 samples. Interviewers then personally contacted the subjects in the chosen secondary sampling units of the 26 through 35 age group and out-of-school 17 year olds. These persons were asked to participate in the assessment.

C. Individuals were classified as black, white, and other on the basis of information provided by the school or by observation. Results were given for black and white only. The number of individuals classified as other was too small to produce reliable results.

IV. Administration of Exercises[11,12]

A. The administration of the exercises was subcontracted to Research Triagle Institute in the East and to Measurement Research Center of Westinghouse Learning Corporation, Iowa City, Iowa, in the West. Cooperation of schools was obtained by first contacting officials at the state and then at school district levels. There was above 90 per cent cooperation by schools.

Adults and out-of-school 17 year olds were contacted by a personal door-to-door household canvass. Each out-of-school participant was contacted individually. The right of each to refuse to participate was respected.

B. A full-time trained staff of 27 district supervisors managed the field work. They were assigned to different geographical areas of the United States. They contacted schools and recruited and trained local teachers to help in the administration of the exercises in schools and recruited and trained other available persons for the out-of-school administration.

C. In the schools, students from a single age group from different classes were brought together in a room for exercise administration. Group size was at least 8 and usually 12 students.

D. The exercises were organized in packages which contained exercises from two or three different subject areas at a single age level. No one person took all the exercises in his age group. Age groups were assessed at different times of the year.

E. In packages administered to groups, taped directions and taped readings of the exercises were used in addition to printed packages. This was done to establish consistency in timing and administration plus to provide for nonreaders.

F. Several packages at ages 9, 13, and 17 consisted of exercises that were given by exercise administrators to one individual at a time. The administration of all the packages for the adult assessment was done by interviews.

---

[11] Womer, op. cit.

[12] Gadway, op. cit.

G. Each package required about 50 minutes of administrative time. Each person took only one package with the exception of the out-of-school, 17 year olds who were asked to take four or five packages each since they were the most difficult and expensive group to locate.

H. Students' names were confidential and did not appear on any packages. The name roster was kept at the building level and used only in the organization of the in-school sampling.

V. Scoring and Analysis[13,14]

A. The scoring and analysis of the exercises were subcontracted to Measurement Research Center of Westinghouse Learning Corporation, Iowa City, Iowa.

B. The multiple-choice exercises were scored and recorded routinely by machine.

C. The openended exercises were scored by trained scorers using a key of acceptable and unacceptable achievements in terms of the objectives.

D. Results were reported for each objective. Also, the results were reported both as the percentage of any group of respondents making the desired responses to an exercise and as the difference between the percentage of a group making the desired responses and the corresponding national percentage.

E. In the assessment, there was a lack of proportionality among characteristics used in the comparison of groups, such as color, sex, parental education. A statistical procedure, balancing, was used to correct for this problem in the comparative analysis of the data. Balancing is a procedure to examine the performance of groups classified on one characteristic adjusting for the fact that these groups differ on a specified set of other characteristics.

VI. Reporting and Dissemination[15,16,17]

---

[13] Womer, op. cit.

[14] Gadway, op. cit.

[15] Womer, op. cit.

[16] Campbell, Report 2, Citizenship: National Results, op. cit.

[17] Gadway, op. cit.

A. The reporting of results was directed to subject matter specialists, professional educators, and informed laymen. Multiple reports were developed to serve these different audiences.

B. Approximately 40 per cent of the exercises were made public at the end of each assessment year. Not all exercises were so reported since they were to be used over again in future assessments in order to measure change by means of comparing the results on the uncontaminated exercises.

C. The exercises released for publication were selected to be representative of all exercises administered as well as the results received on the assessment.

D. Reporting was done by 9, 13, 17, and 26 through 35 age groups. Since the same exercises were used with different age groups, there were comparable data across two or more age levels.

E. Reporting was also done by groups within the categories of regions, community types, sex, socio-economic status, and white, black, and other.

F. Final reports were printed with a short description of the exercises, the national percentage of success, and group differences from the national percentage of success for each exercise. This was done without any interpretation of results.

G. Both observed and balanced results for all exercises and by groups were reported. The effects of balancing on measured characteristics such as sex and region were included in the report.

H. There were no scores reported for individuals. No single individual took more than one twelfth of the exercises, and no individual took a package that sampled only one subject area.

I. Results were reported through the media: written word, radio, television, films, and personal reports.

VII. Utilization of Information[18,19]

---

[18]Womer, op. cit.

[19]Conway, Larry E., "Some Implications of the National Assessment Model and Data for State and Local Education," Paper Presented at the 1973 Annual Meeting of the AERA, New Orleans, Louisiana, February 26, 1973.

A. The results provided potential information for educational decision making. For example, considering the somewhat lower performance of the Southeast Region on the Citizenship results, school boards in that region might decide to conduct their own local investigation to determine the status of their school programs on citizenship skills, understandings, and attitudes.[20]

B. The results raised many questions which may lead to other investigations. For example, in making comparisons of all Citizenship results combined, it was found that the Extreme Affluent Suburbs results were above the national median at all ages and that the Extreme Rural and Extreme Inner City results were below the national median at all ages. Here are discrepancies in performance which need to have causal studies conducted on them from the perspectives of different disciplines such as political science, sociology, economics, and education.[21]

C. The results of several cycles should provide evidence of changes in knowledge, skills, understandings, and attitudes in the age groups as they relate to educational objectives.

D. School administrators can make comparisons between groups, and may improve student performance from the information gained in this manner.

From this review of the model, it is evident that the National Assessment staff has put a great deal of effort and know-how into the design, plus the development of each of the components.

---

[20] Campbell, Vincent N., Manford J. Ferris, and Daryl G. Nichols, National Assessment Report 6, 1969-1970 Citizenship: Group Results for Sex, Region, and Size of Community, Education Commission of the States, National Assessment of Educational Progress, Denver, Colo., July, 1971

[21] Norris, Eleanor L., Vincent N. Campbell, Manford J. Ferris, and Carmen J. Finley, National Assessment Report 9, 1969-1970 Citizenship: Group Results for Parental Education, Color, Size, and Type of Community, Education Commission of the States, National Assessment of Educational Progress, Denver, Colo., May, 1972.

USE OF THE MODEL AT STATE AND LOCAL LEVELS

In planning for the collection of National Assessment data, the model, presented in the last section, was developed. A number of states have found adaptations of the model useful in conducting state assessments in which desirable learning outcomes are identified and the status of learners with respect to these outcomes is determined.

State assessment is a rapidly developing movement. At this writing, all of the states have assessment activities either in operation, in a developmental process, or in a planning stage.[22] While the statewide assessment programs have many similarities, they break down into two basic types of programs on the question, "Who gets to use the results?" The divisions are those states for which data are collected for decision making by state agencies and those states for which data are collected for decision making by teachers and administrators.

In about a third of the states, the programs were mandated by the state legislatures, and the results of the assessments are to be reported back to the state legislatures. In a few of the states, the data are to be used for PPBS (Planning, programming, and Budgeting Systems). In about half of the states where the assessment data are being used to make state-level decisions, state and federal funds will be allocated based on the results. Participation in assessment is required by law in about a fifth of the states. In the states where the assessment data are being used to make

---

[22] State Educational Assessment Programs, 1973 Revision, Joan S. Beers and Paul B. Campbell, "Statewide Educational Assessment," Educational Testing Service, Princeton, New Jersey, 1973, p. 1.

state-level decisions, samples rather than all students are being assessed, while in the local-level, decision-making states all students in the target populations are being assessed. Criterion-referenced instruments are very common with the states where the data are being used for state-level decisions, but the states collecting information for local decision making are favoring norm-referenced instruments. Finally, no dominant funding pattern has evolved in either of the two groups of states.[23]

State Adaptations of the Model

In its assessment of Citizenship education, Maine made an extensive application of the National Assessment model and carefully duplicated it so that comparable data were collected at the state level.[24] Maine's first cycle of the ten subject matter areas of National Assessment (Art, Career and Occupational Development, Citizenship, Literature, Mathematics, Music, Reading, Science, Social Studies, and Writing) is to be completed by scheduling two of these areas each year for five years. Citizenship and Writing were the first subject areas to be assessed.

Based on the results of a previous study of objectives for education in Maine, two review committees decided to accept the National Assessment objectives as being closely related to the Maine objectives. Maine selected the 17-year-old population of in-school students for its first assessment. A sample of 2,000 17-year-old students was used to represent the approximately 17,000 17-year-old students in the State. The State was divided

---

[23] Ibid., pp. 2-3.

[24] Maine Assessment of Educational Progress: Methodology (Report 5), Department of Educational and Cultural Services, Augusta, Maine, 1972.

into four geographical regions. As in National Assessment, school buildings were randomly selected from geographic regions, and students were then randomly selected from the buildings. Packages were developed with exercises taken from the two subject areas. The available, released exercises from National Assessment were carefully examined to see if they reflected objectives valid for Maine and to see if some could be modified, where needed, to be administered in group sessions using the paced-tape method while still retaining a high degree of comparability to the National Assessment individually administered exercises. The packages were made up of 23 Citizenship and seven Writing exercises, plus a 23-item Student Questionnaire. The exercise format was kept virtually identical to the one used in National Assessment. Trained administrators were sent out to administer the exercises, and the exercises were scored according to National Assessment procedures. On data reporting and analysis, there was the census-like reporting of the performance of the Maine students plus comparisons of the Maine results with appropriate National Assessment data.

In summary, the Maine Assessment duplicated the National Assessment procedure as completely as possible. With minor exceptions, the same objectives were used for Citizenship. The same sampling design was used with adaptations to a smaller geographical area and population. The exercises were for the most part taken from those released by National Assessment, and they were organized into packages similar to those used by National Assessment. The administration and scoring of the exercises were conducted in the same manner as National Assessment had used. Since the same private contractors were used by Maine as were used by National Assessment, the duplication was

complete wherever possible. The reporting and data analysis were similar, and the data did provide the opportunity to compare the results in Maine with the results from National Assessment.

Here, the model was very carefully duplicated at the state level. The big question which comes to mind after studying the Maine Citizenship report is, "Aren't the National Assessment data being treated here as some kind of a national norm against which the performances of 17-year-old students in Maine were being compared?" Of course, this use of National Assessment data had been questioned from the start of the proposal for an assessment at the national level. Now, Maine has provided the opportunity to study the effects of this use of the data on the educational system of a state.

Another state which carefully followed the model was Connecticut.[25] Here, an assessment was first conducted in Reading. To permit comparisons, the Connecticut program used available instruments and applicable procedures developed by National Assessment which were adapted to the requirements of the local situation. Connecticut's Reading objectives were matched to the Reading objectives of National Assessment. Approximately 220 reading exercises from National Assessment were used in producing the packages used in the Connecticut assessment. Exercises were selected to represent all of Connecticut's reading objectives. The age groups assessed were 9, 13, and 17. As with the National Assessment packages, tape-recorded instructions were used. The sampling design was a multi-staged design duplicating with few exceptions the National Assessment design. As with National Assessment,

---

[25] Report on the Assessment of Reading Skills of Connecticut Public School Students, Institute for the Study of Inquiring Systems, Philadelphia, Pa., and Department of Education, Hartford, Conn., 1972.

a group of administrators for the packages was recruited and trained.

This was another example of careful duplication of the National Assessment model down to using the same objectives and exercises. Again, there was the use of the National Assessment results as norms to which the Connecticut results were compared.

The Texas Needs Assessment used the model for the development of their assessment in Mathematics at the sixth-grade level.[26] However, while using ideas from the model, they broke with it in a number of places. The Texas people were concerned that the assessment would yield information which would be useful to teachers in their classroom instruction of students. From a pilot study, it was decided to use a criterion-referenced reading test and to work with grade levels instead of age groups of students. They worked with the sixth grade, and the tests were administered by the staff of each school which participated in the assessment. The objectives were chosen from the major skill areas treated in the state-adopted textbooks. Regional location and community size were taken into consideration in selecting the sample. Approximately 10 per cent of the Texas schools teaching at the sixth-grade level administered tests, and approximately 10 per cent of the pupils being taught at the sixth-grade level were included in the sample. Reports were given to teachers on the performance of their individual students. Also, there was a school report on the performance of the students for each school and a report on each of the classes in the school. Comparisons were made on the basis of sex, race, and size of community.

---

[26] Sixth-Grade Mathematics: A Needs Assessment Report, Texas Education Agency, Austin, Texas, 1972.

The Colorado Needs Assessment, while using the model, made an even greater break with it.[27] Its objectives were based on a state study of educational goals, and the educational goals were restated in terms of performance objectives. Following the model, objective-referenced exercises were written. A sampling plan was used and the student responses were analyzed and reported. In this assessment, classroom teachers were involved in the writing and refinement of the behavioral objectives. Objective-referenced exercises were written for nine subject areas and the exercises were administered to a sample of 30,000 Colorado students. A stratified random sampling procedure was used to select a sample of school districts of the State. Then schools were slected at random from the districts chosen. Finally, classes in school buildings were randomly chosen for testing. The samples were representative of all Colorado students in grades 3, 6, 9, and 12. A group of proctors was hired and trained to administer the exercises, and the exercises were scored by computer. The data were analyzed on a statewide and district basis, and the results were broken down by subgroups, e.g. boys, girls, urban, rural.

As pointed out earlier, states are rapidly moving into the assessment field. Some are reproducing the National Assessment model at the state level, and others are developing variations of the model. The more crude efforts have resulted in endless pages of raw percentages without any explanation of the results. Based on a survey of state assessment programs, Beers and Campbell identified several of the problems which are common to these state programs.[28] Naturally,

---

[27] Helper, John W., An Assessment of Learner Needs in Colorado, Colorado Department of Education, Denver, Colo., 1972.

[28] State Educational Assessment Programs, Beers and Campbell, op. cit., p. 3.

a shortage of money and staff were the most frequently mentioned problems, for it is a fact that many states have moved into this area without providing adequate funds for a realistic assessment program. Also, teacher resistance to assessment and negative public attitude toward outside testing were problems mentioned. Test results have been misused in the past, such as the firing of teachers based on incorrect interpretation of test results. Also, test developers have been guilty of violating the privacy of students through questions which transgressed the examinee's human and legal rights. A third problem area has been with the utilization and dissemination of results. Some school officials do not understand the results. In some situations, there has been hostility to the results. Some officials have ignored results in making decisions. Finally, results have frequently not reached the right people in a useable form.

Use of the Model at the District Level

To date, there has been a limited number of efforts reported on the use of the National Assessment model at the district level. Three such assessments on which some data have been released are being conducted in Lincoln, Nebraska; San Bernardino, California; and Montgomery County, Maryland.

In Lincoln, Nebraska, the exercises released by National Assessment in Citizenship and Writing were used in a local assessment which yielded data comparable to National Assessment data.[29,30] A group of supervisory personnel from the central office identified the Citizneship objectives which were applicable to the Lincoln schools.

---

[29] "Weekly Focus," Lincoln Public Schools, Lincoln, Nebraska, February 12-19, 1973, p. 3.

[30] Brandt, Ronald, Associate Superintendent for Instruction, Report on Assessment Results to Board of Education, Lincoln Public Schools, Lincoln, Nebraska, Spring, 1973.

Then the released National Assessment exercises were selected which were applicable to these Lincoln objectives. Also, the National Assessment model was followed in selecting a random sample of 13 year olds from the Lincoln junior high schools. In addition, a sample of in-school, 17 year olds were tested on some of the writing exercises. The administration of the exercises was carried out by a group of specially trained district administrators, and the tape-paced method was used in presenting the exercises to the students. Scoring followed the National Assessment procedures, and in reporting the results comparisons were made to National Assessment data with special attention given to comparable subgroups such as cities of similar size and the same geographic region.

The San Bernardino City Schools developed a criterion-referenced assessment model of student progress which was based on the National Assessment model.[31,32] This model involved local teachers, students, and laymen; eight educational goals were identified through the efforts of workshops involving teachers, students, and patrons. A Curriculum Task Force composed of 20 teachers wrote behavioral objectives for the goals to be appropriate for grades 3, 5, 9, and 12. National Assessment consultants assisted the teachers in developing exercises to assess the stated objectives at these grade levels. Also, the National Assessment consultants helped to design a sampling procedure to provide district-wide representation. The exercises were organized into test batteries for each grade level. The Teacher Task Force administered

---

[31] Bonney, Lewis A., "Application of the National Assessment of Educational Progress Philosophy in San Bernardino City Unified School District," Unpublished Paper, San Bernardino City Unified School District, San Bernardino, California.

[32] Special Curriculum Task Force, "Report on Student Performance," Office of Instructional Services and Research and Development Office, San Bernardino City Unified School District, San Bernardino, California, June, 1972.

and scored the tests. This is a break from the National Assessment practice of using specially trained exercise administrators. The results were tabulated in terms of percentage of students meeting stated behavioral objectives. While the National Assessment model was followed in many ways, such as use of behavioral objectives, criterion-referenced assessment instruments, and sampling of target populations, the assessment was designed for application at the local level, and it was planned, developed, and carried out by local personnel.

The Montgomery County Schools, Maryland, developed a program for assessing 13- and 17-year-old students.[33] In this assessment, the released National Assessment exercises for Writing were used. These were administered in two group-package sessions to samples of 13- and 17-year-old students. The results for Montgomery County students were compared to the results from the nationwide samplings of 13 and 17 year olds by National Assessment. One of the variations in the Montgomery County sampling design was stratification by I.Q. and grade level. The purpose was to spread the sample across the grade by school-I.Q. groups; however, these sampling groups were not used as reporting categories. Each age group (13 and 17) was stratified by I.Q. groups (low and nonlow) and by grade levels.

Adaptation of the Model

In the above discussion, it is evident that there will be as many adaptations of the model as there are local and state units conducting assessments. Probably, there is no specific assessment

---

[33] Bayless, David L., Ralph E. Folsom, and Louise H. Lewis, "Sample Design for Assessing Montgomery County Public Schools 13- and 17-Year-Old Pupils Using the NAEP Model," National Assessment of Educational Progress and Educational Commission of the States, Denver, Colorado, January, 1973.

model which is the best; hence, there is no model that should be applied without modification in any and all situations. Nevertheless, there are principles of good assessment which should be applied in developing or adapting a model for local assessment purposes. Listed below are some characteristics which should be found in a good assessment program.[34]

1. The program has clearly defined goals that apply to a particular audience or audiences.

2. The program has a realistic number of goals which are attainable under the existing assessing conditions.

3. The program has established priorities among its goals and places its major efforts on its major goals.

4. The program has been designed to gather information considered to be important in education.

5. The program has specific objectives which it is striving to attain.

6. The program has been designed to provide results at a useable level of accuracy.

7. The program has used data-gathering instruments which measure the objectives of the assessment.

8. The program has collected data in such a manner as to introduce a minimum of error in the results.

9. The program has scored and processed data in an accurate manner.

10. The program has used analytic techniques that provide the data breakdowns needed by decision makers.

11. The program has reported results in a manner useable by its audience.

12. The program has provided help in the interpretation of results and assistance in their implementation.

13. The program has provided for the active involvement of groups of persons from all of the major audiences for the assessment results.

---

[34] Womer, Frank B., Developing a Large Scale Assessment Program, Cooperative Accountability Project, Denver, Colo., 1973, p. 89.

DISCUSSION OF THE MODEL AS IT APPLIES TO THE CITIZENSHIP ASSESSMENT

As it was developed for the Citizenship Assessment, the model contained several problems. First, ten broad goals were identified as the basis for the National Assessment of Citizenship.[35] These goals were:

1. Show concern for the welfare and dignity of others.
2. Support rights and freedoms of all individuals.
3. Help maintain law and order.
4. Know the main structure and functions of our government.
5. Seek community improvement through active democratic participation.
6. Understand problems of international relations.
7. Support rationality in communication, thought, and actions on social problems.
8. Take responsibility for own personal development and obligations.
9. Help and respect their own families.
10. Nurture the development of their children as future citizens.

These goals were then developed in greater detail through a set of behavioral objectives for each goal. A major question which needs to be asked about the model is, "Do the ten goals cover all of the dimensions of citizenship that we wish to include?" For example, there is no mention in the ten goals of the willingness of the individual to hold public office. While this is a value question, it is possible that there are some serious omissions in this list of

---

[35] Norris, Eleanor L. (Ed.), Citizenship Objectives, Committee on Assessing the Progress of Education, Ann Arbor, Mich., 1969.

ten goals on which the behavioral objectives for the assessment were based.

The behavioral objectives identified for Citizenship were of necessity general in nature. Yet, in a population as large and diverse as that of the United States, there are many subcultures. Undoubtedly, some groups of Americans hold citizenship beliefs which were not reflected in the citizenship objectives which were identified. This could be true for Indians living within a tribal structure on the reservations, and this could be equally true for other groups for whom the statement of behaviors were not acceptable as being descriptive of good citizenship.

Again, the exercises did not describe standards of behavior to which all people in the country could subscribe. The exercises only reflected behaviors which were widely held in the population. One of the obvious difficulties here is that some cultural differences are viewed as undesirable departures from the norm. In other words, if white, Anglo-Saxon protestants from the affluent suburbs responded in a certain manner, then that was considered to be the desirable response, and the responses of other respondents were compared to the so-called desired response. The black population has been especially concerned about this, and they have protested the manner in which their responses have been reported for both the Southeastern part of the country and the core area of the large metropolitan cities. If this is truly a census of the knowledges, attitudes, skills, and understandings of Americans, then the responses of one group should not be negatively compared with the responses of other groups and geographic areas in the country.

Other areas of concern with respect to the model are the limitations of some of the technical procedures used. Based on the broad

goals identified for the Citizenship Assessment, specific objectives were written. The process of writing and screening these objectives included certain difficulties. As stated in the model, this activity was carried out at a center in California. The initial identification of objectives reflected the knowledge and background of the small group of specialists at the center. With this somewhat biased origin, the objectives were then screened by several groups of people who were selected to be more representative of the general population, but a review of the names, addresses, and occupations of these panel members raised questions as to how really representative they were. In addition, National Assessment has very strong political overtones; hence, everything eventually had to be screened for its political implications. Objectives which would be offensive to one group or another in the country were screened out so that the entire operation had, from necessity, a certain blandness about it.

Still another technical difficulty with the model was the transmission problem from one component to the next. One group of specialists developed the objectives and another group wrote the exercises. Hence, the accurate interpretation of the objectives through the exercises was another difficulty. In addition, there was the technical problem of actually writing exercises which truly reflected the behavioral objectives. For some objectives, it was very difficult to write an exercise which would demonstrate if the subject could perform the desired citizenship behavior. Furthermore, the exercise writer might not completely understand the intent of the objective; hence, he failed to present it accurately in the exercises.

Yet another technical problem involved the actual scoring of the exercises. A third group of specialists was hired to score the

exercises. This probably presented little difficulty in the case of the multiple-choice exercises, but with the open-ended kind of exercises, a decision had to be reached as to what was a good answer, a not so good answer, and a poor answer. Then scorers were trained to score the exercises. Naturally, reliability checks were run on the scorers. While it was true that a high level of consistency was achieved with respect to how they scored the exercises, the scoring was still approached from the perspective of a particular group of people who had been trained to take a particular orientation to scoring the responses to the exercises. The perspectives of the various sub-culture groups within our entire society probably were not accurately included within the scorers' perspectives of how the exercises should be scored. This kind of problem always exists in scoring open-ended exercises, and in this situation where a very wide variety of people were included in the sample, undoubtedly many questions were scored from a frame of reference completely alien to that of some of the original responders.

In collecting the data on the out-of-school groups, the administration of the interviews presented still another problem. Admittedly, this was the best procedure for reaching the out-of-school population, and it had the advantage of person-to-person contact. These are very positive traits of the interview procedure; however, the procedure has inherent problems when it comes to reaching members of sub-cultural groups. Those persons who were selected for this work and trained for it undoubtedly possessed characteristics which made them acceptable to the enterprise. On the other hand, these very traits made them unacceptable to certain of the sub-cultural groups with whom they would have to communicate. It is highly probable that the

interviewers were not equally successful with all interviewees and probably were unable to contact certain groups in our population who were hostile to them. This probably had an impact on the sampling results.

Finally, the concept of census-like, value-free reporting is questioned. It is the writer's opinion that the reporting was conducted in as objective a manner as possible, but that at times such things as the organization of the reports, the wording of the reports, and the emphasis in the reporting reflected value positions of the persons at National Assessment who were doing the technical writing. After all, these individuals are all well educated members of the educational establishment. They will filter all of the information reported through their own cultural perspectives in writing the reports. Also, they are sensitive people who are trying to present the information as clearly, honestly, and acceptably as possible. They are going to state things in such a manner as not to be offensive to any particular group, thus there is additional filtering of the information which has implications for what is reported.

In summary, the human element enters into the model at a number of points introducing subjective variables. These are in the selection of goals, the writing of behavioral objectives, the writing of the exercises, the administration of the open-ended interview exercises, the scoring of the open-ended exercises, and the reporting of the data. This is not to say that National Assessment has not succeeded in carrying out as objective an assessment as they could, but it is necessary to point out some of the places where subjective kinds of variables could have been introduced into the model.

DISCUSSION OF THE MODEL FOR STATE AND LOCAL USES

The assessment model has potential for promoting curriculum development. This is especially true when it is applied to state or local situations in the manner used in Colorado and San Bernardino. In these two situations, objectives were developed which specifically applied to the local situation. The statement of well written objectives in behavioral terms may sharpen the purposes of instruction. Through the experience of writing behavioral objectives, the curriculum worker gains a much clearer perception of his task; hence, this practice may have a beneficial impact on curriculum work. On the other hand, the use of behavioral objectives has not always been a positive influence. The objectives may zero in on easily defined behaviors which lack scope and significance. They may produce tunnel vision, and put stress on the inconsequential and trivial. In an effort to be specific and to define the exact behaviors desired, the larger perspective may be lost.

Again, the development of exercises from the identified behavioral objectives may have a positive influence on curriculum. The kind of new, innovative exercises which have been developed by National Assessment may have a very positive influence on what is being taught and how it is being taught. Teachers both in reviewing exercises which have been used in National Assessment and in writing exercises for local assessments may be influenced in their selection of both content and methods by their knowledge of these assessment exercises. Material not relevant to the objectives of the course may be dropped, and methodologies promoting the kind of skills needed in the assessment exercises may be introduced.

On the other hand, the results may be less desirable. If in local and state situations the dictates of finances or the lack of leadership results in the use of poorly written, machine-scored, multiple-choice exercises, the results may be very negative. Teachers may feel pressured to stress rote learning of facts in order to prepare their students for poorly written examinations. Hence, poorly written exercises may keep irrelevant material in the curriculum and limit curriculum innovation and development. The quality of the exercises written and released will have an impact on curriculum development.

Good sampling procedures may give insight into the status of knowledge, understandings, skills, and attitudes of students in a particular target population. This can promote curriculum improvement and innovation. Problem areas in the curriculum may be identified. From the National Assessment, there have been some problem areas identified in the Citizenship results. On an exercise dealing with freedom of speech, a large percentage of 13, 17, and adult age groups indicated that they would not allow sample controversial statements to be made on radio or TV.[36] This showed a lack of understanding or valuing of the Constitutional right of freedom to express controversial or unpopular opinions.

The results on the Citizenship assessment indicated that black, urban students in our large cities compared poorly on knowledge about the structure and function of government to the national average performance on the same exercises.[37]

---

[36] Campbell, Vincent N., et al, Report 2, Citizenship: National Results, Education Commission of the States, National Assessment of Educational Progress, Denver, Colo., November, 1970, p. 34.

[37] Norris, Eleanor L., Vincent N. Campbell, Manford J. Ferris, and Carmen J. Finley, National Assessment Report 9, 1969-1970 Citizenship: Group Results for Parental Education, Color, Size, and Type of Community, Education Commission of the States, National Assessment of Educational Progress, Denver, Colo., May, 1972, pp. 63-65.

On the other hand, there are potential difficulties with assessment data which represent national levels of performance. Even though the data were not collected with this intention and were reported in census-like form, the results of National Assessment are being treated like national norms. Several states have conducted their own assessments duplicating the National Assessment model so that they can make direct comparisons between their state results and the various national, regional, and subgroup results. There is the potential of great mischief in this approach, for it may lead to inaccurate comparisons between groups, states, and regions. In the assessment reports of some states, tables of percentages have been presented without any interpretation or explanation. Some school systems have been presented in a very bad way without any reference being made to the kinds of variables involved in the different learning situations. Such variables as per pupil expenditures, educational level of parents, and motivation of pupils do have an impact on the learning situation. These and other variables should not be ignored in interpreting the results of assessment.

Here, it is not being suggested that assessments should not be conducted because there are potential misuses of the data but those engaged in assessment at national, state, and local levels have the responsibility to be constantly engaged in an educational program to report data in the proper perspective and to aid those using the data to make correct interpretations of it. We need these data for decision making, but if they are misused or misinterpreted, then the decisions based on them may not be good ones.

Finally, where accountability is being applied to a total organization such as a school, a district, or a state, the National Assessment model may be used with little or no modification. It was designed to accurately establish what the level of performance on a given set of objectives was in a population, and it can be used to do this for accountability purposes as well as assessment purposes. Likewise, it can assess subgroups of the population and identify specific strengths or weaknesses in the performance of a given subgroup. The model is an excellent instrument for carrying out accountability in this kind of situation.